



SEEDS

**Working Paper One:
Study Overview
and
Technical Approach**

April 24, 2002

G. Hunolt, SGT, Inc.



Outline

1.0 Introduction

2.0 Roadmap to the Set of Working Papers

Working Paper 1 - Study Overview and Technical Approach

Working Paper 2 - Cost Estimation by Analogy Model

Working Paper 3 - Data Service Provider Reference Model - Functional Areas

Working Paper 4 - Data Service Provider Reference Model - Model Parameters

Working Paper 5 - Data Service Provider Reference Model - Requirements / Levels of Service

Working Paper 6 - ESE Logical Data Service Provider Types

Working Paper 7 - Comparables Database

Working Paper X -References and Acronyms for the Levels of Service / Cost Estimation Working Papers

3.0 LOS and Cost Estimation Study Objectives

4.0 LOS/CE Estimation Study Overview

4.1 LOS/CE Study Tracks

1. Requirements / Levels of Service
2. Cost Estimation by Analogy Model
3. Barkstrom Cost Model
4. Application of COTS Cost Estimation Tools
5. Information Collection - Building the Comparables Database
6. Community Feedback

4.2 Summary Schedule

5.0 Technical Approach

5.1 Notes and Assumptions on the Technical Approach

5.1.1 Data Service Providers, the Reference Model, and Requirements

5.1.2 COTS Cost Estimation Tools

5.1.3 User Model

5.2 Technical Approach - Requirements Analysis

5.3 Technical Approach - Cost Estimation by Analogy Model Development

A.1 References and Acronym List

1 Introduction

This working paper is the first of a set of papers that describes the SEEDS (Strategic Evolution of Earth Science Enterprise Data Systems) Levels of Service (LOS) / Cost Estimation (LOS/CE) study. The study goal is to develop a cost estimation model and coupled requirements and levels of services to support the SEEDS Formulation Team in estimating the life cycle costs of future ESE data service providers and supporting systems, where ‘data service provider’ is used as a generic term for any data/information related activity. The set of working papers is intended to serve as a vehicle for coordinating work on the project, obtaining feedback and guidance from ESDIS SOO and the user community, and as embryos of reports that will be produced as the task proceeds.

As working papers, each version of each paper that appears represents a snapshot in time, with the work in various stages of completion. As work progresses the content (and sometimes the organization) of the working papers will change reflecting progress made, responses to feedback and guidance received, etc.

This first working paper of the set provides an overview of the LOS/CE study, a roadmap to the full set of working papers, a statement of the objectives of the study and an outline of the technical approach being taken to meet the objectives of the levels of service and cost estimation phases of the study. It constitutes a high level plan for the study. The initial version of this paper is focused on the work of the study through June, 2002. A major update to this paper will be provided by June 30, 2002, which will address progress and plans for the next year.

This paper introduces the set of working papers in Section 2. Section 3 states the overall objectives of the LOS/CE Study. Section 4 presents an overview and high level schedule for the study. Section 5 presents the technical approach for the SGT portion of the effort, including some related notes and assumptions, and the approaches to be taken to the requirements analysis and cost model development, consistent with the SGT task plan submitted to ESDIS SOO in October, 2001.

Note: As of late November, 2001, SEEDS replaces the term ‘NewDISS’ under which the Formulation Team had begun work. The new term is intended to emphasize the Earth Science Enterprise’s (ESE’s) evolutionary approach. The term ‘NewDISS’ will be retained when it refers to NewDISS documents that predate the change in terminology. Similarly, the term ‘data service provider’ was adopted as the generic name for an ESE activity that provides any form of data and/or information management and user services, replacing the term ‘data center’, which will be used only in the more conventional sense as a type of data service provider.

2 Roadmap to the Set of Working Papers

This section describes the working papers that together describe the LOS/CE Study.

The initial version of the working papers is a decomposition of the previous overall working paper “SEEDS Requirements / LOS & Cost Model Working Paper - New Year’s Draft”, January 16, 2002, with updates made per the results of the February 5-7, 2002, SEEDS Community Workshop. As an aid to readers of the original document, the roadmap in the initial edition of this paper will contain references to the section numbers in the January 16 working paper from which material was taken.

Although the papers are numbered, except for Working Paper 1 the papers are not intended to be regarded as a sequence, but rather as parallel, and they will refer to each other freely. Their development and updating will be asynchronous, reflecting progress on the study as it occurs.

Working Paper 1 - Project Overview and Technical Approach

The first working paper of the set provides an overview of the SEEDS Levels of Service / Cost Estimation Study, a roadmap to the full set of working papers, and a discussion of the technical approach to the requirements analysis and cost estimation phases of the study, constituting a high level plan for the work to be done. A major update to this working paper will be provided by June 30, 2002. (Contains material from sections 1.1, 3.1, 3.2, 3.3 of January 16, 2002 working paper).

Working Paper 2 - Cost Estimation by Analogy Model

This working paper describes the cost estimation by analogy model that is being developed for this study. This paper will evolve extensively as the work progresses. Its initial focus is on a conceptual description of the model and how it and the cost estimating relationships it uses are expected to develop, scenarios showing how the model will be used, goals and plans for the model prototype planned for June 2002, and the plan for progressively more detailed documentation of computational processes used by the model as it develops. (Contains material from sections 2.1, 2.2, 2.3, 2.4, and 2.5 of the January 16, 2002 working paper.)

Working Paper 3 - Data Service Provider Reference Model - Functional Areas

This working paper describes the concepts involved in the Data Service Provider Reference Model, and describes the functional areas / areas of cost comprising the model. The paper reflects the results of the February, 2002, SEEDS Community Workshop, including drawing on material from white papers submitted by workshop attendees. (Contains material from sections 3.1 and 4.1 of the January 16, 2002 working paper.)

Working Paper 4 - Data Service Provider Reference Model - Model Parameters

This working paper contains definitions of the parameters that are inputs, outputs, and intermediate parameters used by the cost estimation by analogy model, including those that are elements of the comparables database. It constitutes a data dictionary for the model and database. (Contains material from sections 4.2 through 4.5 of the January 16, 2002, working paper.)

Working Paper 5 - Data Service Provider Reference Model - Requirements / Levels of Service

This working paper describes a general set of requirements and levels of service mapped to the functional areas of the Data Service Provider Reference Model. This paper will be maintained and updated as needed through the life of the project. This paper reflects the results of the February, 2002, Community Workshop, draws on white papers submitted by workshop attendees, and includes a new user-oriented view of levels of service. (Contains material from sections 5.1 through 5.11 of the January 16, 2002, working paper.)

Working Paper 6 - ESE Logical Data Service Provider Types

This working paper describes an open set of logical ESE data service provider types, each essentially a group of functions clustered around a different ESE role or mission as an organizing principle. The paper describes how these logical or conceptual provider types relate to physical entities, e.g. real-world data centers that, given their responsibilities within the ESE program, might embody the functionality of several different

provider types. The paper describes how the provider types would be used in ESE architecture studies. The paper reflects the results of the February, 2002, Community Workshop, and draws on white papers submitted by workshop attendees. (Contains material from sections 6.1 - 6.8 of the January 16 working paper.)

Working Paper 7 - Comparables Database

This working paper provides an overview of the Comparables Database, comprising information obtained from existing ESE data activities and other data centers. It includes the database schema or template. It reports on which data centers have provided information to be added to the database, allowing a reader to track the development of the database as the information collection effort proceeds and the paper is updated. The paper does not contain the actual information provided by the sites.

3 LOS and Cost Estimation Study Objectives

Key facets of the SEEDS Formulation study will be to establish the minimum levels of service that ESE data service providers will be required to provide for the user community and to provide the capability to estimate costs for ESE data service providers to provide that level of service. The ultimate objective of the LOS/CE study is to provide the SEEDS Formulation Team with a capability to estimate the cost for various system architectures and mission profiles. Successful development of a life cycle cost estimation capability will be dependent on an accurate assessment of the levels of services needed from ESE data service providers.

“Levels of service” will be associated with certain functional requirements, describing different degrees of performance with which the requirement would be met. For example, a functional requirement might be: “The data service provider shall distribute data and products to users on media”. Accompanying this requirement might be descriptions of quantitatively distinct levels of service, such as “delivery on media shall be provided within one working day of receipt of a data request”, “delivery on media shall be provided within two calendar weeks of receipt of a data request”, and “delivery on media shall be provided within one calendar month of receipt of a data request”. Which level of service would be most appropriate (‘recommended’) or acceptable (‘minimum’) for a particular ESE data service provider would depend on its particular mission and the needs of its users.

The first objective of the LOS/CE study is to assist the Formulation Team in establishing the minimum (and recommended) levels of service (LOS) for ESE data service providers. These LOS will be refined in a bottoms-up manner through community workshops of potential providers and users of ESE data service providers.

The second objective of the LOS/CE study is develop a suite of costs estimation tools that will enable the Formulation Team to estimate the cost impact for various architecture trades, provide NASA Headquarters with estimates of the costs for implementing varying ESE mission profiles and implementation options, and packaging the cost estimation tool kit for use by Earth Science Enterprise scientists responding to new mission opportunities in order for them to estimate the costs for developing and operating the science data ground system for their proposed mission.

The purpose of working on these two objectives together is to ensure that the cost estimation process is tied to a reasonable requirements / levels of service set.

4 LOS / Cost Estimation Study Overview

This section provides an overview of the LOS/CE study. It spans the work being done by GSFC staff, Dr. Bruce Barkstrom of LaRC, as well as SGT.

4.1 LOS/CE Study Tracks

The LOS/CE study is proceeding down a set of parallel tracks, including parallel cost model development efforts. These are seen as a strength, providing an ESE planner or a PI planning a mission with two or three results, and a sense of where and why they differ, will be better grounds for planning a budget than any single estimate. The threads can be held consistent by a common base, the functional areas and at some level the levels of service, and will borrow freely from each other as each refines its description of levels of service at the level of detail appropriate to it. In other words, Dr. Barkstrom's model may operate at a finer level of detail, and so will rely on a more detailed description of levels of service than SGT's model, but the two will operate from a consistent base. Similarly, neither of the cost efforts will contradict the baseline levels of service description produced by the first track with input from and reviewed by the user community.

The LOS/CE Study tracks are:

1. Requirements / Levels of Service

This track involves the development of a baseline set of requirements / levels of service, at a level of detail sufficient to be meaningful to the user, whether a research scientist 'end user' or an intermediate provider or a mission planner or data manager.

The draft levels of service will be updated per input received at the workshop. This will require a second look at the way the information is presented, and at the level of detail presented. The goal will be for the new draft to be as easy as possible for users to evaluate - meaningful but not overly detailed, with levels of service statements with requirements implied, rather than stated explicitly with associated levels of service. User feedback on the LOS draft will be sought.

2. Cost Estimation by Analogy Model

This track involves the development by SGT (Greg Hunolt, Bud Booth) of a cost estimation model using cost estimation by analogy, tied to a set of levels of service and associated requirements spanning all areas of cost at a level of detail consistent with the data available to build the database of comparables the model will rely on, and its inputs and outputs.

Scenarios will be developed to explain the use of the cost model, with the key being to offer the user an unconstrained menu of functions / LOS choices, rather than require that the user select an a priori data service provider type to develop an estimate for. Scenarios will also be developed to explain the use of the cost model to estimate ESE enterprise level costs for alternative architectures of data service provider entities, each embodying one or more logical data service provider types. Feedback on the scenarios will be sought from potential users of the cost model.

3. Barkstrom Cost Model

This track involves a parallel effort (by Dr. Bruce Barkstrom) to produce cost estimates operating at a finer level of detail and employing deeper analytics that will produce a second set of results. Dr. Barkstrom's approach to building a cost model for evolving ESE data systems is:

- a. Start with data life cycle, typically "Prepare, Validate & Produce, Use". "Validate & Produce" activities are dependent on the number of products and the number of versions of each product
Each version goes through four phases: 1) remove blunders, 2) checkout current version, 3) reduce backlog of delayed data, 4) produce data for current stream. "Validate and Produce" activities for both science team and

operations staff are proportional to average number of jobs run per month, with adjustment for degree of automation.

b. Add in data use activities: use Innovation-Diffusion model for spread of data product understanding into the user community. Base staff activity on number of "sales assistance" and "troubleshooting" calls per 1000 user orders. Add in outreach and continuing evolution activities.

c. Collect data missions into organizations, and add in management and infrastructure components of services.

Underlying these three stages, there will be service choices that can be applied to each phase.

For each service choice, there will be technology choices that have both an investment component and a cost component.

This kind of model lends itself to Monte Carlo or simulation approaches, which should make it possible to explore more possibilities than we would otherwise be able to do.

4. Application of COTS Cost Estimation Tools

The LOS/CE study team (David Torrealba, SGT) has completed its survey of commercial-off-the-shelf (COTS) cost estimation tools and recommended acquisition of a suite of tools for estimating lifecycle costs. This suite includes software tools for demand forecasting, neural networks, and case-based reasoning as well as traditional parametric modeling. Using these tools, the study team will investigate potential synergies with cost estimation by analogy and develop a prototype cost model for evolving ESE data systems.

The study team identified three approaches for investigating the potential synergies:

- 1) Non-algorithmic ('machine learning') techniques for estimating costs by analogy
- 2) Demand models based on time series analyses of user access and distribution data
- 3) Function point analysis of 'dataflow centered' models familiar to the science community

After the analogy model (see above) or the machine learning tools (neural networks and case-based reasoning) identify analogies in the 'comparables' database, size data (SLOC or function points) can be applied as inputs to a parametric model. The advantage of this approach is that one will then be able to use Constructive Cost Model (COCOMO) II capabilities to describe and account for differences in development environments. COCOMO II is 'open source' code. Its relationships, algorithms and interfaces are publicly available and well defined. Tools for cost estimation by analogy can be combined with COCOMO II in a relatively straightforward manner.

A 'technical' approach to modeling user demand will employ a standard forecasting tool to perform time series analyses of user access and distribution data. In this case, the investigation will focus on how user demand, if known, may apply to cost estimation models. Part of this work will be to determine the usefulness (for calibrating cost models) of currently available data.

Function point analysis, which is included as the size estimation 'front end' of many parametric models, may support a 'dataflow centered' approach that is more familiar to users in the science community. Function points are the weighted sums of factors that relate to user requirements for data management including numbers of inputs, outputs, logical files, inquiries, and interfaces.

Intended outcomes of continuing this investigation of COTS cost estimation tools are (1) to assist in test and evaluation of a cost model using cost estimation by analogy, (2) to prepare a prototype 'COTS tool kit' to support an analogy model, and (3) to demonstrate a cost model at the Community Workshop in June 2002.

5. Information Collection - Building the Comparables Database

A major effort has begun to collect the information from existing data activities that is needed to build the comparables database. The two step approach that has been adopted for this effort has now begun. A first round of letters has been sent to data centers (by email) asking for documentation or pointers to documentation holding the answers to a set of questions (attached to the letter). The LOS/CE team will attempt to develop the answers to the questions by researching the documentation, and only go back to the data centers for clarifications and to fill gaps.

This process will be refined in response to feedback received during its first round. The area of inquiry (and thus the question list) will be tailored to the intended recipient, and potential recipients will be researched in advance to ensure that only those most germane to the study are asked to participate.

This effort will proceed for many months. The near term intent is to get a sufficient sample to support model development and demonstration of a prototype capability as soon as possible.

6. Community Feedback

The LOS/CE study will not be successful without feedback and guidance from the (at least partially overlapping) community of users and data service providers. The February, 2002, SEEDS Workshop, while representing a first step, generated a number of recommendations about how feedback should be sought. A workshop scheduled for June, 2002, will concentrate on issues of importance to the user community, including how to increase involvement of users in this process, and presenting 'best practices' from organizations internal and external to NASA. Feedback to the model includes (but is not limited to) comments given at workshops, answers to questionnaires, providing white papers for general consumption, being a 'SEEDS prototype', or participating in this study team (for example, by 'tire kicking' model prototypes in the future). Once the model is generated, continued use will enable iterations for improved prediction capability.

4.2 Summary Schedule

February 19, 2002 - Began site information collection effort (continues for a year or more).

March 8, 2002 - Posted Vanessa's workshop results and next steps presentation to the SEEDS website. Follow with the synthesis of workshop results and next steps paper.

March and April, 2002 - Convert February workshop white paper into a set of six smaller white papers, update per workshop results (including a redo of the Levels of Service per the workshop results) and post to SEEDS website.

April, May and June, 2002 - Seek feedback on new LOS draft.

April and May 2002 - Begin building comparables database, adding Benchmark Study data and newly collected site information as received (continues for a year or more).

May and June 2002 - Develop preliminary / placeholder cost estimating relationships for early prototype cost model.

June, 2002 - Report progress on / possibly demonstrate early prototype cost estimation by analogy model (June SEEDS Workshop).

June, 2002 - Report on COTS cost estimating tool survey, discuss most promising tools and how they can be used to support SEEDS cost estimation (June SEEDS Workshop).

The study and schedule will be reviewed in June and this paper will be updated to reflect the results of the review and plans for the balance of FY02 and FY03.

5 Technical Approach

This section outlines the technical approach being taken by SGT to its work to meet the two objectives stated in Section 3 above. The two objectives are inseparably coupled; costs must be driven by requirements, and so the cost estimation tools must be based on a model that maps directly to the requirements set. For this reason SGT's effort consists of two parallel and intertwined paths that will merge in the final product. The first path is a requirements 'levels of service analysis, and the second path is development of a cost estimation capability. The work on the two paths is closely coupled, as the requirements must map to the same framework as the costs, and the concept of a general data service provider reference model (see "Working Paper 3 - Data Service Provider Reference Model - Functional Areas") will be used to provide the common framework.

5.1 Notes and Assumptions on the Technical Approach

This section contains notes that are background for the discussion of the technical approach to the requirements analysis and cost estimation by analogy model development that follow below in Sections 5.2 and 5.3.

5.1.1 Data Service Providers, the Reference Model, and Requirements

The term 'data service provider' is used herein as a broad, generic term for a site or activity that performs all or a subset of the functions defined in the general data service provider reference model. Many well known actual data centers such as the DAACs (Distributed Active Archive Centers) or the NOAA national data centers will perform a subset of the general list of functions, while some sites described as 'data service providers' for this study, e.g. MODAPS (as a sample of a SIPS (Science Investigator-led Processing System), a science team processing facility that does not perform archive or general user distribution), are different in function from many well known data centers but fit within the framework of the data service provider reference model.

The general data service reference model is defined (see Working Paper 3, "Data Service Provider (DSP) Reference Model - Functional Areas") in terms of a set of functional areas, and a set requirements / levels of service is being developed within the functional area framework. These, documented in Working Paper 5, "Data Service Provider Reference Model - Requirements / Levels of Service", will comprise a general set of requirements / levels of service that is independent of any physical entity or architecture. This general set is also a template, in that it contains placeholders for many specifics that would have to be defined in any real case.

The general data service provider reference model will have subsets corresponding to the tentatively defined ESE logical data service provider types (see "Working Paper 6 - ESE Logical Data Service Provider Types", which discusses the current open set of types), seen as logical functional groupings based on an ESE role or mission as an organizing principle. This approach has the advantage of allowing the future definition of additional data service provider types, or variations of the types defined herein, i.e. other possible subsets, within the framework of the general model. In the same manner, the general set of requirements / levels of service will have subsets corresponding to each of the defined data service provider types.

Data services provider types, as logical groups of functions, do not correspond to physical entities (e.g. data centers or flight project data systems) except in a case where a physical entity performs a single ESE role or mission that corresponds to a data service provider type. In most cases, physical entities / organizations will have more complex ESE roles and missions that would correspond to some combination of the same or different logical data service provider types.

As in the case of the overall general set of requirements / levels of service, the requirements / levels of service set for a data service provider type will be also be a template containing placeholders for quantitative

parameters that would be defined for a specific instance of a data service provider of that type. For example, suppose that a cost estimate is needed for a simple case where an entirely new organization is being set up to perform the functions of a single data service provider type. A requirement in the template for that type might be that “the data service provider shall provide an archive capacity of [number TB]”. If the actual mission of the data service provider required that it archive certain data streams and generated products that would accumulate to a total volume of 100 TB, then that value would be inserted into the template, with the result being a specific requirement for that data service provider (i.e., “the data service provider shall provide an archive capacity of 100 TB”) that could then be used in the process of generating a cost estimate for the data service provider.

5.1.2 COTS Cost Estimation Tools

The use of COTS cost estimation tools is being explored, for example for software development, to check the results of the cost estimation by analogy model, to provide alternate results for evaluation, or perhaps to replace it for aspects of the cost modeling where a COTS tool proves to be superior in tests against the independent cases. This requires an examination and evaluation of the available COTS tools, selecting the most promising for test, and exercising them. The SGT report “Survey of Cost Estimation Tools, Final Report” by David Torealba, February 28, 2002, reports on progress in this area.

5.1.3 User Model

The life cycle cost model will need to project user demand for a data service provider’s services over a period of time. In addition to data service provider history information, the effort will include an examination of existing user models including Dr. Bruce Barkstrom’s.

5.2 Technical Approach - Requirements Analysis

The requirements / levels of service developed by this study are intended to support the cost modeling effort, and not to serve as the complete definition of the requirements side of a contract between the SEEDS program office and ESE data service providers, or as a basis for procurements. The requirements will be ‘end-to-end’ in that they will encompass all significant elements of cost, and will be directly and explicitly traceable to cost.

The requirements analysis would proceed as follows:

- a. Review existing NewDISS and ESE program documents, and incorporate the draft “NewDISS Level 0 Requirements, September 2001”, as a high level programmatic framework. Review the EOSDIS Level 2 Requirements for Version 0 as a reference to a previously defined set of requirements and levels of service that could be a source for the current effort. Review USGCRP and CES reports (see references) for additional input on requirements / levels of service.
- b. Develop an initial general requirements / levels of service template corresponding to an initial version of a general data service provider reference model, and subsets corresponding to an initial set of logical data service provider types, consistent with program documents, for review, revision as needed, and approval as a starting point by the SEEDS Formulation team.
- c. Use a community workshop to get user feedback on, and input into, the requirements and levels of service definitions. Produce updates to the requirements templates set, for review, revision as needed, and approval by the SEEDS Formulation team.
- d. Work with the user and provider community and the SEEDS Formulation Team to obtain feedback on and guidance for the improvement of the requirements / levels of service.
- e. Produce a final requirements templates set for review, revision as needed, and approval by the SEEDS Formulation team, and produce a final report on the requirements analysis. (SGT has a contract deliverable for a final set of requirements / levels of service on March 31, 2002, but will maintain and update the document as work progresses over the life of the task.)

At each step, as changes to the requirements sets are approved by the Formulation team, ensure that the requirements changes are reflected back into the data service provider reference model.

5.3 Technical Approach - Cost Estimation by Analogy Model Development

The cost estimation model will be based on a ‘comparables’ or ‘cost by analogy’ method; it will estimate costs using cost estimating relationships derived from a number of existing data service providers that are functionally comparable to the different types or combinations of different types of ESE data service providers.

The general approach is to draw on the data service provider reference model concept developed for the Best Practices / Benchmark study and to develop a cost estimation model that estimates the cost of an ESE data service provider based on the actual costs of comparable data service providers, a “cost estimation by analogy” methodology. Information about other ESE or outside data service providers will be collected to provide the best possible basis for comparison. The cost estimating relationships that will be used by the model will themselves be developed and evolve as the comparables database is built (see “Working Paper 2 - Cost Estimation by Analogy Model” for a description of how this process is seen).

The life cycle cost model will project user demand for a data service provider's services over a period of time. In addition to data service provider history information, the effort will include an examination of existing user models including Bruce Barkstrom's.

The cost estimation by analogy model development will proceed with the following steps:

- a. Define, and refine based on feedback from ESDIS and the Formulation Team, the content of a data service provider cost estimate; i.e. what elements of cost at what level of detail with what supporting information are required as the output product from the cost estimation tool. The further development of the cost model would be guided by the results that the model must produce, allowing for the fact that this will change as the effort proceeds.
- b. Survey available COTS cost estimation tools, evaluate and test the tools that seem most likely to be useful for this study, and produce a report summarizing the results of the survey and recommending tool(s) to be used further in the study.
- c. Obtain and examine the Bruce Barkstrom user model and any other user model that might be useful for this study.
- d. Extend the existing Best Practices / Benchmark study reference model to encompass the full range of data service provider functions, refine the original list of model parameters, add implementation, parameters necessary for estimation of cost, etc. (See “Working Paper 4 - Data Service Provider Reference Model - Model Parameters” for definition of the model’s parameter set.)
- e. Derive subsets of the general reference model to correspond to the logical data service provider types. These subsets will include the functional areas and metrics appropriate to each data service provider type.
- f. Map the information collected on selected data service providers during the Best Practices / Benchmark study to the extended reference model to begin to build the model’s comparables database;
- g. Identify additional data service providers to be added to the comparables database. Draw from DAACs not included in the Best Practices / Benchmark study, SIPs, selected ESIPS. Consult with ESDIS to arrive at a list of candidates. Reserve some data service provider cases for use as independent test cases for the cost model.
- h. Collect the additional or update information and add to the model’s information set (i.e., as was done for the Best Practices / Benchmark study, map data service provider information to the reference model’s common set of metrics).

Note that steps b and c can run in parallel with a, d, etc. Also, steps g and h can run in parallel with d, e, and f provided that information collected early on can be supplemented as completion of steps d, e, and f identify gaps in the initial collection.

- i. Use the mapped data service provider information to construct relationships (for each data service provider type, within each functional area) relating actual data service provider staffing and costs and known development effort and workload performed, etc. These relationships are currently TBD but could include linear regression equations and the like. Probable errors of estimate will also be derived for each relationship.
- j. Test the model by inputting information for the independent test cases and determining the degree to which model is able to correctly calculate staffing, costs, etc. Test the COTS cost estimation tools, to determine which should be incorporated into the model or used in conjunction with the model to give the best possible overall result. Also consider incorporation of the Bruce Barkstrom or other externally developed user model.
- k. Obtain community feedback on the prototype cost estimation model.
- l. Release a life-cycle cost model Version 0 that incorporates initial user feedback; continue to obtain and incorporate community feedback by presenting study results and providing prototype models for hands-on peer review.
- m. Provide a final report, and provide cost estimates for ESE data service providers as needed.

A.1 References and Acronym List:

The References Section and the Acronym List for all of these Working Papers is in the document “References and Acronyms for the Levels of Service / Cost Estimation Working Papers ”.